Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method comprising:

extracting a set of uniform resource locators (URLs) from ~~at least~~ one document or from multiple documents associated with a single web host;

identifying sub-strings occurring in multiple URLs in the set of URLs as session identifiers, based on a particular rule and based on the sub-strings occurring in multiple URLs of the set of URLs;

~~analyzing the set of URLs extracted from the at least one document to determine those in the set of URLs that contain session identifiers by locating the session identifiers in the set of URLs extracted as sub-strings that occur in multiple URLs of a web site;~~

generating a clean set of URLs from the set of URLs ~~extracted from the at least one document~~ by removing the session identifiers; and

determining when at least one particular URL has already been crawled based~~, at least in part,~~ on a comparison of the particular URL to the clean set of URLs.

2. (Cancelled)

3. (Currently amended) The method of claim 1, ~~wherein~~ where the ~~at least one~~ document or each of the multiple documents is a web document downloaded from a web site.

4. (Currently amended) The method of claim 1, ~~wherein~~ where the comparison of the particular URL to the clean set of URLs comprises calculating a fingerprint value for a particular URL and for each of the URLs in the clean set of URLs, and where the comparison is based on a comparison of [[a]] the fingerprint value of the particular URL to the fingerprint values of the URLs in the clean set of URLs ~~calculated for each of the URLs in the clean set of URLs~~.

5. (Currently amended) The method of claim 1, ~~wherein~~ where the particular rule comprises:

~~the session identifiers are determined as including~~ determining that the sub-strings ~~from the set of URLs that~~ do not reference content.

6. (Cancelled)

7. (Currently amended) The method of claim 1, ~~wherein the analyzing the set of URLs extracted from the at least one document further includes~~ where the particular rule comprises:

~~locating the session identifiers in the extracted set of URLs as~~ determining that the sub-strings [[that]] contain characters consistent with a session identifier.

8. (Previously presented) The method of claim 1, further comprising:

downloading content from the particular URL when the particular URL is determined to not already have been crawled.

9. (Currently amended) The method of claim 1, further comprising:

storing information based on the clean set of URLs for use in later determining whether additional URLs have already been extracted; and

storing the set of URLs ~~extracted from the at least one document~~, including embedded session identifiers, for use in later accessing the set of URLs ~~extracted from the at least one document~~.

10. (Currently amended) A method comprising:

receiving a set of uniform resource locators (URLs);

analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing the set of URLs to identify ~~those~~ one of the sub-strings as corresponding to a session ~~identifiers~~ identifier based on multiple occurrences of [[a]] the sub-string in the set of URLs.

11. (Currently amended) The method of claim 10, ~~wherein~~ where the set of URLs are extracted from a web document associated with a web host.

12. (Currently amended) The method of claim 10, ~~wherein~~ where the set of URLs are extracted from multiple web documents associated with a single web host.

13. (Currently amended) The method of claim 10, further comprising:

removing identified session identifiers from the set of URLs; and

storing the set of URLs, with the removed session identifiers, as a clean set of URLs.

14. (Previously presented) The method of claim 13, further comprising:

adding a generated session identifier to URLs in the clean set of URLs.

15. (Currently amended) A device comprising:

at least one fetch bot configured to download content on a network from locations specified by uniform resource locators (URLs);

a content manager configured to

extract URLs from the downloaded content, and

identify session identifiers from the URLs extracted from the downloaded content based, at least in part, on multiple occurrences of the session identifiers from a single web site; and

a URL manager configured to create clean versions of the URLs extracted from the downloaded content by removing the session identifiers from the URLs and to store the clean versions of the URLs ~~extracted from the downloaded content in which the session identifiers are removed from the URLs extracted from the downloaded content~~.

16. (Currently amended) The device of claim 15, ~~wherein~~ where the content manager is further configured to identify the session identifiers based on locating sub-strings, within the URLs extracted from the downloaded content, that contain characters consistent with session identifiers.

17. (Original) The device of claim 15, further comprising:

a database configured to store the downloaded content.

18. (Currently amended) The device of claim 15, ~~wherein~~ where the URL manager is further configured to determine when additional URLs have previously been stored by comparing clean versions of the additional URLs to the stored clean versions of the URLs extracted from the downloaded content.

19. (Currently amended) The device of claim 15, ~~wherein~~ where the session identifiers include characters from the URLs extracted from the downloaded content that do not reference content.

20. (Currently amended) A device comprising:

means for receiving a set of uniform resource locators (URLs);

means for analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

means for further analyzing the set of URLs to identify ~~those~~ one of the sub-strings as corresponding to a session ~~identifiers~~ identifier based on multiple occurrences of [[a]] the sub-string in the set of URLs.

21. (Currently amended) The device of claim 20, ~~wherein~~ where the set of URLs are extracted from a web document associated with a web host.

22. (Currently amended) The device of claim 20, ~~wherein~~ where the set of URLs are extracted from multiple web documents associated with a single web host.

23. (Original) The device of claim 20, further comprising:

means for removing the identified session identifiers from the set of URLs; and

means for storing the set of URLs with the removed session identifiers as a clean set of URLs.

24. (Previously presented) The device of claim 23, further comprising:

means for adding a generated session identifier to URLs in the clean set of URLs.

25. (Currently amended) A computer-readable ~~medium~~ memory device including programming instructions that when executed by at least one processor causes the at least one processor to perform a method including:

receiving a set of uniform resource locators (URLs);

analyzing the set of URLs for sub-strings that are structured in a manner consistent with session identifiers; and

further analyzing the set of URLs to identify ~~those~~ one of the sub-strings as corresponding to a session ~~identifiers~~ identifier based on multiple occurrences of [[a]] the sub-string in the set of URLs.

26. (Currently amended) The computer-readable ~~medium~~ memory device of claim 25, ~~wherein~~ where the set of URLs are extracted from a web document associated with a web host.

27. (Currently amended) The computer-readable ~~medium~~ memory device of claim 25, ~~wherein~~ where the set of URLs are extracted from multiple web documents associated with a single web host.

28. (Currently amended) The computer-readable ~~medium~~ memory device of claim 25, ~~wherein~~ where the programming instructions further include programming instructions that cause the at least one processor to:

remove the session identifiers from the set of URLs; and

store the set of URLs with the removed session identifiers as a clean set of URLs.

29. (Currently amended) The computer-readable ~~medium~~ memory device of claim 28, ~~wherein~~ where the programming instructions further include programming instructions that cause the at least one processor to:

add a generated session identifier to URLs in the clean set of URLs when the URLs are to be used to access a web document.